Cancer Genome Biology at the Broad Institute: A “Team of Teams”

Levi A. Garraway, M.D., Ph.D.
Overarching Goals of Broad Cancer Genome Characterization Efforts

• A complete catalogue of significant and impactful tumor genomic alterations
• To address major questions in cancer biology using genomics
• Clinical applications of genome sequencing data
Platforms Leveraged by Broad Research Teams

- Biological Samples Platform
- Chemical Biology Platform
- Genome Sequencing Platform
- Genetic Analysis Platform
- Imaging Platform
- Metabolite Profiling Platform
- Proteomics Platform
- RNAi Platform
Cancer Genome Characterization at the Broad Institute: The “Core” Team

- 3 Senior Associate Members (faculty)
- 4 Associate Members (faculty)
- 4-6 Research Scientists
- >20 Computational Biologists
- >20 Postdoctoral fellows/students
- Many technicians, project managers, software engineers, etc.
- Many collaborators
Characteristics of Cancer Team Science Projects at the Broad Institute

• Many cancer genome projects, large & small
  – TCGA (GCC and GDAC)
  – NHGRI Sequencing center-initiated projects
  – Broad faculty-driven initiatives
  – Collaborator-driven initiatives
  – Academic-industry collaborations (e.g., CCLE)
  – “Clinical sequencing” projects
  – Philanthropic projects

• Considerable breadth and diversity of genomic data
  – Whole genome, whole exome, “targeted” exome, transcriptome (“RNA-seq”), methylome...
Cancer Genome Sequencing Process Flow

Sample Intake
- Compliance Review
- Quantification
- Quality Check

Genotype Characterization
- Fingerprinting

Illumina Sequencing
- Whole Exome
- Whole Genome
- Custom Targeted
- RNAseq

Data QC
- Preliminary SNPs
- Concordance
- Indel cleaning
- dbSNP%

Integration with other omic data

Validation/extension

Experimental follow-up

Findings that may impact cancer biology or clinical oncology

“Firehose” pipeline
- Base mutations
- Insertions/deletions
- Copy number alterations
- Rearrangements
- Pathgens

Experimental clinical oncology

Experimental follow-up
Cancer Genome Projects: Specific Hurdles to Overcome

• Process oversight and “de-mystification”
  – Who controls the queue/timetable?

• “Lost in the ether”
  – What happened to my samples/data?

• Production-level “admixture”
  – WGS on Monday; WES on Tuesday, RNA-seq on Wed...

• Bureaucratic and logistical delays
  – Shifting consent form criteria, personnel absences, etc.

• Managing computational bandwidth
  – 1 TB per 60X T/N pair (whole genome)!

• Efficiency of mutation validation/extension
  – We need 200 more T/N pairs now!

• Publication/authorship considerations
  – Who gets to be 1st and last on this 60+ author paper?
Cancer Genome Projects: The “Operational Unit”

- A triad of “champions” owns each project
- Disease biology: often postdocs, grad students
- Analytical: both post- or pre-PhD with supervision
- Presumes dual first and senior authorship
Cancer Genome Projects: The “Operational Unit”
Cancer Genome Projects: The “Operational Unit”
Phasing increases efficiency of personnel utilization.

~months
The Cancer Genome Steering Committee

• 4 faculty, 6 staff scientists (2 co-chairs)
• Strategic >> operational guidance
• Scientific input at project “pivot points”
  – (e.g., experimental plan after a key genomic insight is made)
• Identification of systematic errors/issues
• Resource management
Cumulative Cancer Samples Sequenced at Broad

Next-Generation Sequencing Pipeline

Picard pipeline
Broad's Sequencing Platform

Calibrate quality scores

Align to genome (MAQ, BWA)

Mark duplicate reads

BAM file

Visualization (IGV)

dbGAP

QC

Mutations

Indels

Purity ploidy

Copy-number

Rearrangements

Fusion-genes

Pathogens

Annotation + Reports

muTect

Indelocator

ABSOLUTE

SegSeq

dRanger

PathSeq
Output (past ~2-3 years)

• ~60 papers (several others submitted/in press)
• >5000 cancer genomes
• Multiple new sources of funding
  – NHGRI Sequencing center grant renewal
  – U01 in Exploratory Clinical Sequencing (with help from NCI)
  – Other R01, U01, P01, R33 grants
  – Multiple foundation grants
  – Industry-sponsored research
  – Other philanthropy
Framework for vetting and prioritizing the next wave of large-scale projects?

- Compelling scientific rationale
- Potential for high impact (scientific or clinical)
- Deeply invested collaborator(s)
- Local “champion”
- Technical feasibility (e.g., the samples are ready to go, the consent form is kosher, protocol active)
- Validation/follow-up plan
- Funding is in place
# Flagship Project Concept: (Prostate Cancer, early 2011)

<table>
<thead>
<tr>
<th>Relevance: biological insights of importance?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Timely, opportunity</td>
<td>Yes</td>
</tr>
<tr>
<td>Address key biological or clinical questions</td>
<td>Indolent versus lethal disease; resistance to antiandrogen therapies, relationship between somatic genomics patterns and ancestry</td>
</tr>
<tr>
<td>Systems in place for such questions</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study Design: comprehensive in breadth or depth</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to expand to multi-dimensional genomics</td>
<td>Yes</td>
</tr>
<tr>
<td>Discovery cohort in place</td>
<td>Nearly 300 samples in place, most are frozen tissue</td>
</tr>
<tr>
<td>availability / access to extension cohort</td>
<td>Extensive collaborative network in place, both FFPE, frozen tissues and derivative cells</td>
</tr>
<tr>
<td>Model system – comparative oncogenomics</td>
<td>GEMM systems that model leading genetic/biological drivers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Follow-through: Coordinated efforts /collaborations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional validations</td>
<td>Yes, active and ongoing</td>
</tr>
<tr>
<td>Model systems</td>
<td>Yes, established and emerging ones</td>
</tr>
<tr>
<td>Path to translation</td>
<td>Extensive translational / clinical-trial investigators engaged</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Logistics: funding, staffing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Funding for genomic discovery</td>
<td>CIP</td>
</tr>
<tr>
<td>Funding for extension studies</td>
<td>Sources available (PCF, Movember, DOD...)</td>
</tr>
<tr>
<td>Funding for downstream studies</td>
<td>Funded Starr, DOD grants, SPORE application likely, PCF</td>
</tr>
<tr>
<td>Faculty champions</td>
<td>Levi Garraway</td>
</tr>
<tr>
<td>Biology champions</td>
<td>Sylvan Baca</td>
</tr>
<tr>
<td>Analysis champions</td>
<td>Mike Lawrence</td>
</tr>
<tr>
<td>Disease experts collaborators</td>
<td>Kantoff/Rubin/Tewari/Balk/Bubley/Taplin</td>
</tr>
</tbody>
</table>
Broad Cancer Genome Sequencing: Lessons Learned for Team Science

• Deep and sustained collaborations are essential
• All parties must “buy-in” and receive due credit
• “Ground level” ownership by nimble teams
• Data generation is the easy part!
• Think like a biologist, act like a CEO/COO
• “Hub” model for team science research?
• High-level team science cannot happen everywhere